Evaluation of the psychometric properties of the Icelandic version of the CORE-OM, its transdiagnostic utility and cross cultural validation

Short title: Psychometric properties and transdiagnostic utility of the Icelandic CORE-OM

Hafrún Kristjánsdóttir,

University of Iceland/ Reykjavik University/Landspítali-The National University Hospital of Iceland,

Address: Menntavegur 1

Tel: +3548941713

Fax: +3545995201

E-Mail: hafrunkr@ru.is

Baldur Heiðar Sigurðsson,

Landspítali-The National University Hospital of Iceland/Reykjavik University (baldurhs@landspitali.is),

Paul Salkovskis,

University of Bath (P.M.Salkovskis@bath.ac.uk),

Daníel Ólason,

University of Iceland (dto@hi.is),

Engilbert Sigurdsson,

University of Iceland/Landspítali-The National University Hospital of Iceland/Reykjavik University (engilbs@landspitali.is),

Chris Evans,

Nottingham University (chris@psyctc.org),

Eva Dögg Gylfadóttir,
University of Iceland (edg1@hi.is)
and

Jón Friðrik Sigurðsson,

University of Iceland/Landspítali-The National University Hospital of Iceland/Reykjavik University (jonfsig@landspitali.is).

Corresponding author: Hafrún Kristjánsdóttir, Landspítali-The National University Hospital of Iceland/University of Iceland/Reykjavík University

This research was sponsored by: Landspítali – University Hospital Science fund.

University of Iceland research fund and Wyeth research found.

Abstract

Background: Development of transdiagnostic standardised measures of psychological distress have contributed to the development of practice based evidence networks. The translation and validation of such measures cross culturally is important if such research is to be generalised across health care systems in different countries.

Method: Translation of the Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM) from English into Icelandic was undertaken according to recommended protocols. The resulting instrument was evaluated for reliability and validity in three groups; patients undergoing psychological treatment in general practice (n= 289), psychiatric outpatients (n=98) and student controls (n=207). Clinical participants underwent diagnostic interview were the MINI-International Neuropsychiatric Interview were administered. The clinical participant also completed the CORE-OM and the Beck Depression and Anxiety Inventories before and after treatment. The transdiagnostic relationship between CORE-OM scores and scores on those diagnostic measures was assessed.

Results*:* Good levels of validity, reliability and internal consistency were found for the CORE-OM and its domains and sensitivity to change over treatment shown. As anticipated, the Risk domain had different characteristics than other domains. CORE-OM scores correlated strongly with both BDI-II and BAI, particularly the problem domain.

Conclusion: The Icelandic translation of the CORE-OM is psychometrically sound and can be applied in Icelandic mental health studies as it has been in English speaking settings. Nevertheless the validity of the Icelandic version of the CORE-OM needs to further investigate in larger and more diverse samples

Key words:

CORE-OM, Psychometric properties, Transdiagnostic

Key Practitioner Message

•The psychometric properties of the Icelandic version of CORE-OM are comparable to the original English version

•The results indicate transdiagnostic utility of the CORE-OM

•The CORE-OM is a valuable instrument in outcome research on psychological treatment, specifically transdiagnostic treatment

Introduction

A lack of consensus on which measures to use for any particular construct complicates reviews of outcome research since scores have to be transformed for comparability and, it is not always clear that such transformations are psychometrically valid. Both researchers and clinicians have been concerned about this for several decades and since the 1970's various attempts have been made to develop a core measure that clinicians can agree is reliable, valid and appropriate across common emotional disorders(Horowitcz, Strupp, Lambert & Elkin,1997)

These attempts have not been without difficulties, one being that theoretical neutrality of the measure was considered a condition for a widespread consensus but choice of items tends to be linked to the theoretical stance of the author/s. However, progress was achieved in the 1990's when a number of assessment instruments were developed, such as the Outcome Questionnaire system (OQ) (Lambert, Hansen & Harmon, 2010) Treatment Outcome Package (TOP)(Kraus & Castonguay, 2010) and the CORE system(Barkham et al., 1998; Evans et al., 2000; Evans et al., 2002). Of particular

importance is the way in which the development and validation of such systems opened the way to the evaluation of studies that were either not diagnostically driven or spanned multiple diagnoses. Such therapies have often been labelled transdiagnostic therapies and these approaches to mental health problems have gained increasing interest in resent years. In such approaches the specific diagnosis is seen as less important and the specific cognitions, emotions and phenomenology the clients bring are more the focus of the therapy(Harvey, Watkins, Mansell, & Shafran, 2004).

The CORE System consist of CORE-OM (a self-report outcome measure, long and short version) and CORE-A (a clinician rated assessment sheet filled out before and after treatment). CORE-OM is the focus of this paper and described in more detail in the measurements section. It was designed to be pantheoretical and not to be diagnosis specific. Studies have reported acceptability, reliability, validity in terms of clinical/non-clinical differentiation and sensitivity to change. Convergent validity with diagnosis specific measures and other non-specific measures has been shown and a score translation to and from the Becks Depression Inventory – second edition (BDI-II) has been established for the English version(Evans et al., 2002). These parameters are explored for the Icelandic translation in this paper.

However, no study we know of has explored the transdiagnostic issue of how the overall score and domain scores of the CORE-OM relate to different diagnostic measures in a multiple linear regression or partial correlation analysis. Such analyses show us (i) whether the diagnosis specific measure have equal and distinguishable covariance with the non-diagnostic measure or if one accounts for more of the shared

variance than the other; and (ii) whether covariance is confined to the problem domain on the CORE-OM or more general. Extremely low levels of unique variance explained for one of the instruments might (although not conclusively) suggest that the CORE-OM covers a wide ranges of symptoms common across different diagnoses which can be said to render it with transdiagnostic features. Such a measure would be appropriate as a primary measure in research on transdiagnostic treatment and even as secondary measure in research on a wide range of diagnosis specific treatments. Given such results it may be further hypothesised that the problem domain would be very important for such transdiagnostic features since the domain is designed to be equally symptom oriented as the BDI-II and the Becks Anxiety Inventory (BAI). That could be tested in a similar but reverse fashion, i.e. by assessing which domain score explains the most variance of the BDI-II and the BAI, and then how much unique variance the other domains explain. These transdiagnostic features have been intrinsically assumed, but have to our knowledge not been investigated in this way yet.

The psychometric properties of the CORE-OM in English in UK data have been shown to be excellent(Connell et al., 2007; Evans et al., 2002; Lyne, Barrett, Evans, & Barkham, 2006), internal consistency is high (Cronbach's alpha of .94 for overall score in both non-clinical and clinical samples), and test retest stability of .90 in a student sample again for overall score(Evans et al., 2002). The validity of CORE–OM in terms of "caseness" is very good as is its discriminant validity shown in a large location difference between clinical and non-clinical samples. Results also showed good convergent validity and the CORE–OM correlates most highly with conceptually close measures, e.g. the BDI–II, (r = 0.81)(Leach et al., 2006) and the SCL–90–R, (r = 0.88).

The CORE-OM is also sensitive to change and capable of detecting clinical change. Finally, the CORE-OM is well accepted by patients(Evans et al., 2002) and free to reproduce on paper without licence fees.

The measure in English has been evaluated in a range of settings in the UK and South Africa(Campbell & Young 2011). Barkham, Culverwell, Spindler,Twigg and Connell (2005) found that the psychometric properties of the CORE-OM in an older adult UK population were acceptable and overall similar to the psychometric properties of the instrument in the general population in UK though clinical means were lower in the older adult sample than in younger clinical samples.

Barkham et al (2001) also provided benchmark data, i.e. baseline scores from 39 secondary healthcare services, to which other service profiles can be compared. UK National benchmarks and a list of over 100 publications that have used the system can be accessed on the CORE system's web site, www.coreims.co.uk.

The CORE-OM has been now translated into 20 languages, Norwegian, Welsh, Spanish, Portuguese, German, Dutch, Greek, Italian, Danish, Icelandic, Swedish, Polish, Finnish, Lithuanian, Slovak, Turkish, Croatian, Albanian and Gujarati. The psychometric properties of the Italian, Slovak, Swedish, Lithuanian, Japanese and Portugalete translations have been published to date. The psychometric properties of the Italian version of the CORE-OM were good(Palmieri et al., 2009). The same was true for the Swedish version (Elfström, Evans, Lundgern, Johansson, Hakeberg, & Lundgern, 2012) as well as the The Slovak version (Gampe, Bieščad, Balúnová-

Labaničová, Timul'ák, & Evans, 2007), Japanies version (Uji, Sakamoto, Adachi & Kitamura, 2011) and the Lithuanian version (Viliū nienė et al, 2012). The Interal reliability in in the Portugalish version of CORE –OM was good in Sales, Molerio, Evans and Alves (2011) study.

The Icelandic version of CORE-OM

The idea of practice research networks (PRN) in other countries, using the same primary measure that is free to reproduce, is an attractive one and would support empirical exploration of practice similarities and differences between countries. The team that developed the CORE system had that in mind when they defined necessary features for such a measure, and they envisioned that it would be translated to other languages(Barkham et al., 1998; Evans, et al., 2000). They therefore established some rules for the translation process to ensure comparability between translations. Those ground rules are evolving but involve: a) multiple forward translations, b) holding a focus group comparing those translations with a member of the original author group present in the room and answering questions about the intention behind aspects of the measure or about the English, c) the translators should work with at least one mental healthcare professional, and d) input should be obtained from at least one service user who speaks both English and the target language. The Icelandic version of the CORE system was translated according to these ground rules and in close collaboration with one representative from the CORE system (CE). It was translated by a team of clinical psychologists and psychiatrists in addition to a professional translator.

The present study reports psychometric data on the Icelandic version of the CORE-OM, specifically internal reliability, test-retest reliability, convergent validity, discriminant validity and sensitivity to change. Particular attention is paid to relationships with measures of anxiety and of depression to throw light on the transdiagnostic aspirations of the measure, which to our knowledge has not been explored in this way before.

Method

Participants

The study included 594 adult participants comprising three samples, one non-clinical sample and two clinical samples. The non-clinical sample consisted of 207 university students from two different department, department of law and department of psychology. Oft the 207 students, 71 (34%) were males and 136 (66%) females, with mean ages 22.9 and 22.5 years respectively. The students had no knowledge about CORE-OM prior to the administration All students were approached twice during lecture two weeks apart. There were no exclusion criteria for this sample.

The two clinical samples consist of primary care patients and psychiatric out-patients, respectively, both experiencing psychiatric problems, primarily mood and anxiety disorders. Frequencies of psychiatric problems, evaluated with The MINI-International Neuropsychiatric Interview (see more information in measurement selection), in the two clinical samples can be seen in table 1. All participants in the clinical samples attended 5 week transdiagnostic cognitive behavioral group therapy. The group therapy was carried out as a part of standard care both in primary care and out-patients ward at the university hospital. Exclusion criteria for treatment were age under 18 years,

current psychotic symptoms, current self-reported substance use dependence or obvious signs of dementia or other generalized cognitive impairment.

The primary care sample consists of 289 patients, 41 (14%) males and 248 (86%) females, with mean ages 41.2 and 40.3 years, respectively. All participants in the primary care sample were referred to the study by general practitioner. The psychiatric ward sample consists of 98 patients, 29 (30%) males and 69 (70%) females, with mean ages 35.4 and 35.3 years, respectively. All participants in the psychiatric ward sample were referred to the study by medical doctors, psychiatrists, psychologist or psychiatric nurses

Of the total sample, 577 (97%) participants returned appropriately completed questionnaires and were included in the study. The return for each sample was; 204 (99%) university students, 279 (96%) primary care patients and 94 (96%) secondary care patients. Of the 204 students who completed the questionnaires satisfactorily the first time, 153 (74%) returned completed questionnaires two weeks later for test retest analysis. Of the 363 patients who completed the questionnaires satisfactorily at the first therapy session, 230 (63%) returned completed questionnaires at the end of treatment for sensitivity to change analysis.

Measures

Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM)(Evans et al., 2000; Evans et al., 2002) is a 34 item self-report instrument, with four domain scores, each reflecting a different aspect of life: well-being (four items),

problems/symptoms (twelve items), life functioning (twelve items) and risk to self and others (six items). Items are scored on a five-point scale from 0 = not at all to 4 = all the time. Total scores range between 0 and 4 and higher scores reflect more severe problems.

Beck Depression Inventory-II (BDI-II)(Beck, Steer & Brown, 1996) is a 21 item self-report instrument used to measure depressive symptomatology. The twenty-one symptoms of depression are rated on a four-point scale (0-3), within the time frame of the past two weeks. The inventory had earlier been translated into Icelandic back translated to ensure accuracy. Psychometric evaluation of the Icelandic version of BDI-II revealed adequate alpha coefficients in both student and patient sample (α = .91 and α=.93 respectively) and further confirmatory factor analysis supported the unidimensionality of the scale (Arnarson, Olason, Smári, & Sigurdsson, 2008).

Beck Anxiety Inventory (BAI)(Beck, Epstein, Brown, & Steer, 1988) is a 21 item instrument designed to evaluate symptoms of clinical anxiety experienced during the past week. The items are rated on a four point scale (0-3). The instrument had earlier been translated into Iceland and back translated. A psychometric analysis showed that the Icelandic version of the BAI has acceptable psychometric properties(Sæmundsson et al., 2011) with alpha coefficients in both student and patient samples, i.e. α = .96 and α=.92, respectively.

The MINI-International Neuropsychiatric Interview (MINI)(Sheehan et al., 1998) is a short structured diagnostic interview of mental disorders according to the diagnostic criteria of the DSM-IV and ICD-10. MINI was designed for multicentre clinical trials

and epidemiological studies. Icelandic version of MINI has not yet been extensively studied although one preliminary study gives some support to its validity (Sigurðsson, 2008). The English version of MINI has shown excellent reliability (Lecrubier et al., 1997). In Lecrubier et.al study (1997) kappa coefficients as well as sensitivity and specificity were good or very good for all diagnosis except GAD, bulimia and agoraphobia. Inter-rater reliability and test retest reliability were also good in the same study. Table 1 lists MINI diagnostic categories and frequencies for each category.

Table 1 about here

Procedure

The CORE-OM, the BDI-II and the BAI were administered to the students (Sample 1) during class hours and again two weeks later. An effort was made to make sure that there was a sufficient space provided between students to ensure privacy during responding. Students were not paid for their participation. The same measures were administered by clinical psychologist to the clinical samples (Samples 2 and 3) in intake interviews prior to the transdiagnostic CBT group therapy and again at the end of therapy. The MINI was only administered to the clinical samples. The therapy was developed by experienced clinical psychologists with extensive experience in delivering cognitive behavioural therapy. The treatment was delivered by two qualified clinical psychologists in each group, once a week for two hours for five weeks.

Written informed consent for participation was obtained from all participants. Permission for the study was obtained from the National Bioethics Committee in

Iceland (VSNb2005090003/03-15) and the study was approved by the Icelandic Data Protection Authority (S2602/2005).

Results

Age and gender

Using Independent sample t - test there were no significant main effects on CORE-OM for gender in all samples (Students: t (202) = 1.76, $p > 0.05$. CI: -0,06 –to 0,23; Primary care: t (278) = -1.10, $p > 0.05$ CI: -0,34 to 0,10; Secondary care: t (92) = 0.38, $p > 0.05$, *CI:*-0,23 to 0,34). The results indicate that non gender cut off scores should be used. The effects of age were explored. Scattergrams showed no evidence of non-linear relationships and Pearson coefficients were calculated to test for simple correlation. Correlations were r = 0.06 in the secondary care sample, r = 0.25 in the primary care sample and r = 0.09 in the student sample. Age therefore accounts for only 6% of the CORE-OM variance in the primary care sample and less than 1% in the other samples. Student sample may suffer from restricted range. The results indicate that it is not necessary to adjust scores on the basis of either gender or age for the CORE-OM in Iceland.

Reliability

Internal consistency

A measure cannot have validity unless it has reliability. Table 2 shows the Cronbach's alpha coefficients for all the CORE-OM scores. The alpha coefficients for the CORE-OM total score are satisfactory in all samples. Those for the domain scores are also

acceptable (range 0.78 – 0.95) for all the CORE-OM domain scores although being lower for the Risk domain (0.64 – 0.73).

Table 2 about here

Test – retest reliability

In total 206 students were approached and 153 (74%) returned completed questionnaires from both administrations. In order for the results to be comparable to Evans et al. (2002), Spearman's rhos were calculated. Test–retest correlations were high and acceptable for the CORE-OM Total score (rho = 0.80) and for all the CORE-OM domain scores (rho = 0. 75 - 0.78) except for the Risk domain (rho = 0.49). Low test-retest on the Risk domain is not surprising when the distribution of scores on the domain is examined: reassuringly, of the 153 participants 127 scored zero on the pre-test and 120 scored zero on the post-test. While such skewed results are expected for a non-clinical sample, they mean that small changes by a few participants have a large impact on stability parameters.

Since the CORE-OM scores can be dichotomised using clinical/non-clinical cutting points (as described below), Kappa coefficients for test-retest reliability of the dichotomised scores were calculated and ranged from 0.61 – 0.71 for the total score and all the domain scores except for Risk (κ = 0.46). These values are categorised as "substantial" agreement for all but Risk, where agreement is moderate according to the Landis and Koch criteria (Landis & Koch, 1977). The Kappa and rho values are shown in Table 3.

Table 3 about here

Validity

Signal Detection Analysis / ROC Analysis

Signal Detection Analysis (SDA) was conducted in the pooled clinical sample in order to map the CORE-OM scores to the MINI diagnosis. A ROC curve was obtained for all the scores and the total score. The Youden Index (J)(Youden, 1950) was used to determine the optimal cut-off values from the ROC analysis choosing the criterion score that maximises J. As another way to estimate the optimal cut-off value, the scores in non-clinical sample were compared to scores in the pooled clinical samples, using the Jacobson & Truax method C(Jacobson & Truax,1991) to estimate the cut-offs for clinical significant change (CSC). The results are summarized in Table 4.

Table 4 about here

The Risk domain was not included in the analysis because reliability analysis suggests that it is not a reliable scale (although it serves a clinical flag) and is therefore not suitable for a ROC analysis that i assumes a spectrum is measured against a dichotomous measure . Therefore it is wise to use the CORE-OM non-Risk total domain which consists of the total score minus the score for the Risk items to predict diagnosis on the MINI. The scores on the CORE Total scale are much lower than the non-Risk

domain. When the two methods are compared it can be seen that the criteria obtained by the CSC and ROC analysis are similar for the non-Risk domain, but not for the CORE Total score where the criteria obtained with CSC are lower than the criteria according toYouden's J.

Convergent validity

Pearson´s product moment correlation coefficients were computed between the CORE-OM domain scores and the BDI-II and the BAI (see Table 5). Across the domain scores, correlations were highest between the conceptually related measures, i.e. the CORE Well-being, Problems and Functioning scores and the BDI-II and the BAI, showing acceptable convergent validity. The CORE-OM total score correlated strongly with the BDI-II (r=0.88) and somewhat less with the BAI (r=0.72). The CORE-OM domain scores other than the Risk score correlated strongly (between 0.81 and 0.85) with the BDI-II and somewhat less with BAI (between 0.63 and 0.74). The correlations between the Risk score and BAI and BDI-II were low (0.36 and 0.55 respectively). When the BDI-II was correlated with the CORE-OM controlling for the BAI, the correlation was 0.79, but for the BAI controlling for the BDI-II, the correlation dropped much more, being 0.33.

Table 5 about here

These partial correlations suggest that the BDI-II accounts for much of the CORE-OM variation indicating that the CORE-OM measures features nonspecific to diagnostic categories. These results was further explored further by comparing two regression models; the first one entering only the BDI-II and the second one adding BAI to measure the unique additional explanatory power of BAI when controlled for BDI-II (planned stepwise regression). The measures were entered into the model in this order on grounds of the partial correlations above. Table 6 shows the results, which indicate that the BDI-II explains almost 80% of the variance of the CORE-OM total score. The

additional explanation of the BAI is also significant, but only accounting for a further 2%. Results are virtually the same for the domain scores and non-risk items total score. Because of this, the instrument that explained most of the variance of the CORE-OM, namely BDI-II, was used as the only reference instrument in all further analysis.

Table 6 about here

In order to examine the differential predictive power of individual domains of the CORE-OM for BDI-II, a stepwise linear regression was computed with the BDI-II as the dependent variable and the CORE-OM scores as independent variables. The results are presented in Table 7. The Table shows that the problem domain, which was designed to measure symptoms and complaints of those who suffer from depression and anxiety, covers almost all the variation and the other scores add little to the explained variance though each added a statistically significant covariation sequentially. The CORE-OM scores explained a total of 79% of the variance on the BDI-II, there of the Problem score explained 71%, indicating that the Problem domain has by far the strongest covariance with the BDI-II, the other domain scores adding only a further 9% though the gain through adding each was statistically significant at $p<.001$.

Table 7 about here

Discriminant validity

The means and standard deviations for the CORE-OM domain scores are shown in Table 8. Mean scores were highest in the secondary care sample, followed by the primary care sample on all domain scores and the Total score. Lowest mean scores on

the CORE–OM were in the non–clinical sample as expected. A one way Analysis of Variance (ANOVA) indicated significant main effects for groups on all the scores and the total score (F ranging from 15.39 to 97.31). In order to examine further the differences between the samples, a series of post hoc tests were conducted (with Tamhane corrections for unequal variances as there was significant heterogeneity of variance). There were significant differences between the non-clinical sample and both clinical samples for all the CORE-OM scores and the CORE-OM Total scale ($p<0.05$). The differences between the two clinical samples were however only found on the total scale and the Problem score ($p<0.05$).

Table 8 about here

Sensitivity to change

A repeated measures ANOVA was carried out to assess sensitivity to change for non-risk total score. This was a 2x2 analysis; the patient group (primary care and secondary care) was the grouping variable and the pre-post comparison the within subject factor. There were significant main effects of repeats ($F[1,241]=33.84$, $p<0.001$) but no interaction between group and repeats ($F[1, 241]=1.47$, $p=0.23$). This suggests that the CORE-OM non-risk total scores changed during treatment in both samples that indicates it´s utility as a primary measure in transdiagnostic treatment. Table 9 shows descriptive statistics and effect sizes for the domain scores and the total scores.

Table 9 about here

Discussion

This paper presents the psychometric properties of the Icelandic version of the CORE-OM. Our aim was to establish the reliability and validity of the Icelandic translation, to assess its sensitivity to change and to suggest appropriate cut-off scores. A secondary aim was to look into the possible utility of the instrument as a primary measure in transdiganostic treatment.

The psychometric properties of the Icelandic version of the CORE-OM are as good as those found for the English version in UK data (Evans et al., 2002). As in the UK data, the internal consistency and test-retest reliability were found to be good for both the non-clinical and the clinical samples, with the exception of the Risk domain score. These results are consistent with the CORE-OM results of Evans et al.(2002). The six items covering the Risk domain include four items addressing the risk of harming yourself and two items addressing the risk of harming another person. These two risk factors are not necessarily related which further reduces the internal reliability already constrained by having so few items.

Test-retest reliability was good, though slightly lower than in the Evans et al.(2002) study. The Icelandic version of the CORE-OM proved to be meaningfully stable over two week time interval in a substantial non-clinical sample. One step not considered for the UK data was to use Kappa to explore the stability of a "clinical/non-clinical" dichotomisation of scores. This too showed satisfactory reliability.

As expected, the British cut off scores did not prove appropriate for use with the Icelandic translation. The total score and two out of four domain scores have lower cut off scores according to Youden's J for the Icelandic data than the British cut offs calculated by CSC criterion. When that criterion is applied to the Icelandic data all domain scores and the total score have lower cut-off scores than in the 2002 British data(Evans et al., 2002). Firstly, this may reflect cultural differences between Icelandic people and British in how they present their mental health problems. Secondly, it may also reflect differences in demographic variation between the countries. The Icelandic population is so far more homogeneous regarding e.g. ethnic background and socioeconomic status, although in the last decade or so Iceland has become more and more multicultural. Thirdly, recent Icelandic epidemiological surveys of psychological health and morbidity are lacking and there may be a possibility that diagnostic base-rates differ between the countries and different cut-off scores reflect that. Finally, these differences may reflect different selection biases in the clinical samples since in Iceland people are not steered in different groups based on level of adjustment as in the British samples.

However, the present study yielded a cut off score for the total score almost identical to the 2007 British data(Connell et al., 2007). In addition the cut-off scores for men and women in all three samples were almost identical in the Icelandic data which suggest that the same cut-off scores should be used for both genders. Connell et al.(2007) also recommended the use of a single cut off score, although the gender difference was still slightly larger than in the Icelandic data.

Convergent validity of the Icelandic version of the CORE-OM appeared to be good with high correlations between it and the conceptually related measures, BDI-II and BAI. These results are highly consistent with the findings of Evans et al.(Evans et al., 2002) and suggest that the convergent validity of the Icelandic version of the CORE-OM is at least as good, as that of the original British version. High inter-correlations were found between all the domains except the Risk domain in all three samples. The Risk domain had a slightly lower correlation with other domains in the student sample than in the clinical samples. Overall these findings are again similar to those of Evans et al. (2002) and Lyne et al.(2006) apart from the Risk domain, which had slightly higher correlations with other domains in the clinical samples than in student sample. A multiple regression analysis resulted in the BAI explaining close to nothing of the CORE-OM total score and all domain score variance, not already explained by the BDI-II.

As predicted the Problems domain accounts for the majority of the variance of the BDI-II with much less variance being accounted for by the other domains. This is likely explained by the fact that the BDI-II is highly oriented towards the experience of mood problems and depressive symptoms. This is further underscored by its higher sensitivity than all other domain scores, including the non-risk items score and the total score. Any differential convergent validity for the Function, Well-being and Risk score validities will need testing by correlation against more specific measures such as quality of life (e.g., QOLS(Burckhardt, Anderson, Archenholtz, & Hägg, 2003) and work and social adjustment scales(Marks,1986). The validity of the Risk domain may be best assessed against the measurement of hopelessness and suicidal risk (e.g., the BHS and the BSI scales(Beck, Brown, Berchick, Stewart & Steer, 1990). Further research is needed

though strong differential validity of the domain scores was never a design intention for the CORE-OM as noted above.

There was a clear difference between the non-clinical and the clinical samples on all the scales and the total score. The differences between the two clinical samples showed that the secondary care sample scored significantly higher on the Total and the Problem scores, but not on the other scores, indicating that the CORE-OM differentiates acceptably between patients with mental health problems and the general population. The failure to detect differences on most domain scores between the two clinical samples may relate to the way services operate in Iceland. Everyone across the country has direct access to the Mental Health Services at Landspitali-The National University Hospital of Iceland in Reykjavik, irrespective of whether they have been referred by their GP or not. In fact, patients have a choice where they seek mental health services and they are therefore not necessarily steered into the service by defined criteria. It is therefore probably unsurprising that no differences were observed between the primary and the secondary care samples. In keeping with this the analysis of sensitivity to change, comparing the primary and the secondary care samples, there was no evidence of either a group effect or a group X domain interaction, suggesting that the samples are more similar than different. The Icelandic version of the CORE-OM appears to be sensitive to changes following treatment. However there was no control group so it's not certain that the changes are due to treatment rather than time.

The similarity of our results to those of Evans et al.(2002) suggests that the Icelandic translation of the CORE–OM was successful and that Icelandic studies using the

CORE-OM will be fully comparable with studies using the original British version. This will offer international comparison between treatment studies in Iceland and Britain. Moreover, the instrument is sensitive to change in a diagnostically heterogeneous sample. These results may seem to contrast with the results of the hierarchical regression analysis suggesting that the CORE-OM scores are well explained by the BDI-II with little added explanatory value of the BAI. This could indicate that the CORE-OM is mostly a measure of depression. However, the results may even more likely reflect the functional relationship between different emotional problems. A person with long lasting anxiety may suffer from decreased adjustment resulting in depressive symptoms and vise versa. A high correlation between measures of anxiety and depression are therefore to be expected and hence little additive explanatory value of variables added in later stages in hierarchical regression analysis such as those carried out here. These results taken together therefore suggest, at least tentatively, that the CORE-OM could be a feasible choice as a primary measure in efficacy studies of transdiagnostic treatment.

Acknowledgements

The authors are grateful to the participants for their time and patience and the psychologists who ran the treatment groups and administered the psychological instruments.

References

Arnarson, T. O., Olason, D. T., Smári, J., & Sigurdsson, J. F. (2008). The Beck Depression Inventory Second Edition (BDI-II): psychometric properties in Icelandic student and patient populations. *Nordic journal of psychiatry*, *62*(5), 360–365. doi:10.1080/08039480801962681

Barkham, M, Culverwell, A., Spindler, K., Twigg, E. & Connel, J (2005). The CORE-OM in an older adult population: psychometric status, acceptability, and feasibility. *Aging & mental health*, *9*(3), 235–245. doi:10.1080/13607860500090052

Barkham, M, Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., … McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: toward practice-based evidence in the psychological therapies. Clinical Outcomes in Routine Evaluation-Outcome Measures. *Journal of consulting and clinical psychology*, *69*(2), 184–196.

Barkham, Michael, Evans, C., Margison, F., McGrath, G., Mellor-Clark, J., & Milne, D. (1998). The rationale for developing and outcome batteriesfor routine use in service settings and psychotherapy outcome research implementing core. *Journal of Mental Health*, *7*(1), 35–47. doi:10.1080/09638239818328

Beck, A T, Brown, G., Berchick, R. J., Stewart, B. L., & Steer, R. A. (1990). Relationship between hopelessness and ultimate suicide: a replication with psychiatric outpatients. *The American journal of psychiatry*, *147*(2), 190–195.

Beck, A. T, Steer, R. A., & Brown, G.K. (1996). *BDI-II, Beck Depression Inventory II: Manual (2nd ed.).* The Psychological Corporation, Harcourt, Brace, and Company: Boston.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, *56*(6), 893–897. doi:10.1037/0022-006X.56.6.893

Burckhardt, C. S., Anderson, K. L., Archenholtz, B., & Hägg, O. (2003). The Flanagan Quality Of Life Scale: evidence of construct validity. *Health and quality of life outcomes*, *1*, 59. doi:10.1186/1477-7525-1-59

Campbell, M. M, & Young, C. (2011). Introducing the CORE-OM in a South African context : validation of the CORE-OM using a South African student population sample. *South African Journal of Psychology*, *41*(4), 488–502.

Connell, J., Barkham, M., Stiles, W. B., Twigg, E., Singleton, N., Evans, O. & Miles, J. N. V. (2007). Distribution of CORE-OM scores in a general population, clinical cut-off points and comparison with the CIS-R. *The British journal of psychiatry*,*190*, 69–74. doi:10.1192/bjp.bp.105.017657

Elfstöm, M.L., Evans, C., Lundgren, J., Johansson, B, Hakeberg, M & Carlsson, S.G. Validation of the Swedish Version of the Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM). *Clinical Psychology & Psychotherapy.publish online 22 march 2012.* DOI: 10.1002/cpp.17

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: psychometric properties and utility of the CORE—OM. *The British Journal of Psychiatry*, *180*(1), 51–60. doi:10.1192/bjp.180.1.51

Evans C., Mellor-Clark J., Margison F., Barkham M., Audin K., Connell J., & McGrath G. (2000). CORE: Clinical Outcomes in Routine Evaluation. *Journal of Mental Health*, *9*(3), 247–255.

Gampe, K., Bieščad, M., Balúnová-Labaničová, L, Timul'ák, L, & Evans, C. (2007). Slovak adaptation of CORE-OM (Clinical Outcomes in Routine Evaluation - Outcome Measure) | Slovenská adaptácia metódy CORE-OM [Abstract]. *Ceska a Slovenska Psychiatrie*, *103*(1), 4 – 13.

Harvey, A., Watkins, E., Mansell, W., & Shafran, R. (2004). *Cognitive Behavioural Processes across Psychological Disorders: A Transdiagnostic Approach to Research and Treatment* (1st ed.). Oxford University Press, USA.

Horowitcz, L.M., Strupp, H.H., Lambert, M.J., & Elkin, I. (1997). Overview and summary of the core battery conference. In Strupp, M.J., Lambert, M.J., & Horowitz, L.M (Eds.), *Measuring Patient Changes in Mood, Anxiety and Personaility Disorders: Toward a Core Battery* (pp. 11 – 54). Washington DC: American Psychological Association.

Jacobson, N, & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12 – 19.

Kraus, D, & Castonguay, L. (2010). Treatment Outcome Package (TOP) – Development and use in naturalistic settings. In Barkham, M, Hardy, G.E., & Mellor-Clark, J. (Eds.), *Developing and Delivering Practice Based Evidence – A Guide for the Psychological Therapies* (pp. 155–174). Chichester, UK; Malden, MA.: Wiley-Blackwell.

Lambert, M.J., Hansen, N.B., & Harmon, S.C. (2010). *Outcome Questionnaire system (The OQ System): Development and practical applications in healthcare settings*. Chichester, UK: Wiley-Blackwell.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Leach, C., Lucock, M., Barkham, M., Stiles, W. B., Noble, R., & Iveson, S. (2006). Transforming between Beck Depression Inventory and CORE-OM scores in routine clinical practice. *The British journal of clinical psychology / the British Psychological Society*, *45*(Pt 2), 153–166. doi:10.1348/014466505X35335

Lecrubier, Y, Sheehan, D.V., Weiler, E, Amorim, P., Bonora, I, & Harnett, K. (1997). The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. *European Journal of Psychiatry*, *12*, 224 – 231.

Lyne, K. J., Barrett, P., Evans, C., & Barkham, M. (2006). Dimensions of variation on the CORE-OM. *The British journal of clinical psychology / the British Psychological Society*, *45*(Pt 2), 185–203.

Marks, I.M. (1986). *Behavioural Psychotherapy: Maudsley pocket book of clinical management.* Bristol: John Wright.

Palmieri, G., Evans, C., Hansen, V., Brancaleoni, G., Ferrari, S., Porcelli, P., … Rigatelli, M. (2009). Validation of the Italian version of the Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM). *Clinical psychology & psychotherapy*, *16*(5), 444–449. doi:10.1002/cpp.646

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., … Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview

(M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of clinical psychiatry*, *59 Suppl 20*, 22–33;quiz 34–57.

Sigurðsson, Baldur Heiðar. (2008). *Comparison between two standardised psychiatric interviews and two self-report measures: MINI, CIDI, PHQ and DASS* (Cand Psych). University of Iceland.

Sæmundsson, B. R., Þórsdóttir, F., Kristjánsdóttir, H., Ólason, D. Þ., Smári, J., & Sigurðsson, J. F. (2011). Psychometric properties of the Icelandic version of the Beck Anxiety Inventory in a clinical and a student population. *European Journal of Psychological Assessment*, *27*(2), 133–141. doi:10.1027/1015-5759/a000059

Uji, M., Sakamoto, A., Adachi, K. & Kitamura, T. (2011). Psychometric properties of the Japanese version of the Clinical Outcomes in Routine Evaluation-Outcome Measure. *Comprehensive Psychiatry,53(5),* 600-608. DOI:10.1016/j.comppsych.2011.09.006

Viliū nienė R, Evans C, Hilbig J, Pakalniškienė V, Danilevičiū tė V, Laurinaitis E & Navickas A. (2012). Translating the Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM) into Lithuanian. *Nordic Journal of Psychiatry.* Posted online on December 11, 2012. (doi:10.3109/08039488.2012.745599

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32–35.

Table 1. MINI diagnostic categories and frequencies in pooled clinical sample.

Disorder	Time frame	Frequencies(%)
Major depressive episode	Current (2 weeks), Recurrent	45 30
MDE with melancholic features	Current (2 weeks)	19
Dysthymia	Current (2 years)	18
Suicidality	Current (past month)	44
Manic episode	Current, Past	2
Hypomanic episode	Current, Past	2
Panic	Current (past month), Lifetime	8
Agoraphobia	Current	3
Social phobia	Current (past month)	31
OCD	Current (past month)	7
PTSD	Current (past month)	7
Alcohol dependence	Past 12 months	8
Alcohol abuse	Past 12 months	1
Substance dependence (Non alcohol)	Past 12 months	2
Substance abuse (Non alcohol)	Past 12 months	0
Psychotic	Current Lifetime	0
Mood disorder with psychotic features	Current	1
Anorexia nervosa	Current (past 3 months)	0
Bulimia nervosa	Current (past 3 months)	2
GAD	Current (past 6 months)	44
Antisocial personality disorder	Lifetime	3

Table 2. Internal reliability (Cronbach α) of the CORE-OM scales.

	Non-Clinical sample		Primary Care		Secondary Care		Pooled clinical samples	
Scale	Alpha	CI(95%)	Alpha	CI(95%)	Alpha	CI(95%)	Alpha	CI(95%)
Well-being	0.84	0.82 - 0.86	0.79	0.75 - 0.83	0.78	0.70 - 0.85	0.79	0.75 - 0.82
Problem/symptoms	0.91	0.90 - 0.92	0.87	0.84 - 0.89	0.88	0.84 - 0.91	0.87	0.85 - 0.89
Functioning	0.88	0.87 - 0.90	0.88	0.85 - 0.90	0.84	0.79 - 0.87	0.87	0.85 - 0.89
Risk	0.67	0.62 - 0.71	0.64	0.57 - 0.70	0.73	0.63 - 0.81	0.66	0.61 - 0.71
Non-risk items	0.95	0.95 - 0.96	0.94	0.93 - 0.95	0.93	0.91 - 0.95	0.94	0.93 - 0.95
All items	0.95	0.94 - 0.96	0.94	0.93 - 0.95	0.93	0.91 - 0.95	0.94	0.93 - 0.95

Table 3. Test-retest reliability of the CORE-OM scales.

	Present study					Evans et al. (2002)
	N	Kappa	CI(95%)	Spearmans rho	CI(95%)	Spearmans rho (n =43)
Well-being	152	0,63	0.47 -0.79	0,75	0.67 - 0.81	0,88
Problem	148	0,64	0.48 - 0.80	0,77	0.69 - 0.82	0,87
Function	149	0,61	0.44 - 0.78	0,75	0.67 - 0.81	0,87
Risk	153	0,46	0.26 - 0.65	0,48	0.35 - 0.60	0,64
Non-risk items	151	0,63	0.42 - 0.84	0,71	0.62 - 0.78	0,91
All items	148	0,71	0.59 - 0.83	0,8	0.73 - 0.84	0,9

Table 4. Results of the Signal Detection Analysis and CSC.

	ROC				CSC		
	AUC	Criterion	Sensitivity %	Specificity %	Criterion	Sensitivity %	Specificity %
Problems	0.80	1.42	72.1	78.7	1.33	80.1	64.0
Well-being	0.79	1.75	61.1	81.3	1.35	74.2	64.0
Function	0.83	1.33	63.1	86.7	1.08	77.4	69.3
Non-risk items	0.84	1.25	75.8	77.3	1.21	77.7	74.7
All Items	0.83	0.62	81.7	70.3	0.98	54.0	86.5
BAI	0.75	9,45	77.0	64.0	10.83	72.3	66.7
BDI-II	0.82	19.50	63,4	86,7	12.66	84.8	57.3

.

Table 5. Correlations between the CORE-OM scales and the other self-report instruments.

	BAI	Well being	Problems	Functioning	Risk	All items	Non-risk items
Non-clinical sample							
BDI-II	.68**	.83**	.84**	.82**	.52**	.89**	.76**
BAI		.64**	.72**	.61**	.38**	.70**	.63**
Well Being			.85**	.81**	.48**	.91**	.82**
Problems				.80**	.49**	.95**	.86**
Functioning					.48**	.93**	.84**
Risk						.59**	.09*
All items							.86**
Primary care sample							
BDI-II	.55**	.78**	.77**	.79**	.51**	.85**	.82**
BAI		.51**	.67**	.53**	.35**	.63**	.63**
Well Being			.79**	.77**	.50**	.89**	.87**
Problems				.75**	.50**	.93**	.93**
Functioning					.41**	.57**	.54**
Risk						.54**	.58**
All items							.98**
Secondary care sample							
BDI-II	.58**	.75**	.78**	.76**	.53**	.84**	.82**
BAI		.53**	.59**	.47**	.31**	.58**	.56**
Well Being			.85**	.73**	.41**	.90**	.88**
Problems				.72**	.46**	.94**	.94**
Functioning					.40**	.90**	.87**
Risk						.47**	.49**
All items							.98**
pooled clinical sample							
BDI-II	.55**	.78**	.78**	.79**	.52**	.85**	.82**
BAI		.51**	.64**	.51**	.34**	.61**	.61**
Well Being			.81**	.76**	.48**	.89**	.88**
Problems				.74**	.49**	.94**	.93**
Functioning					.35**	.54**	.51**
Risk						.52**	.56**
All items							.98**

Table 6. BDI-II and BAI explained variance of CORE-OM total scale and domain scores.

	R^2-change	B	T
Well-being			
BDI-II	0.69	0.06	22.61*
BAI	0.01	0.01	4.51*
Problems			
BDI-II	0.72	0.05	22.44*
BAI	0.04	0.02	9.37*
Function			
BDI-II	0.68	0.05	22.70*
BAI	0.01	0.01	3.24*
Risk			
BDI-II	0.26	0.01	8.92*
BAI	0.00	0.00	1.49
Non-risk Items			
BDI-II	0.60	0.03	17.69*
BAI	0.02	0.01	5.24*
All items			
BDI-II	0.79	0.04	28.90*
BAI	0.02	0.01	7.46*

*p< 0.01

Table 7. CORE-OM domain scores incremental explained variance of BDI-II.

	R^2change	B	T
BDI			
Problems	0.71	4.79	8.43*
Function	0.06	4.99	8.57*
Well-being	0.02	3.01	6.01*
Risk	0.01	2.86	3.66*

*p<0.001.

Table 8. Descriptive statistics and ANOVA results for sample differences for the CORE-OM scales.

	Non-clinical sample *n* = 204	Primary care *n* = 279	Secondary care *n* = 94		Pooled clinical sample *n* = 373	Effect size clinical/non-clinical
Domain	Mean (SD)	Mean (SD)	Mean (SD)	F-value*	Mean (SD)	
Well-being	0.89 (0.78)	1.89 (0.21)	2.11 (0.97)	89.808	1.94 (0.99)	0,11
Problems	0.91 (0.68)	1.76 (0.82)	2.03 (0.82)	97.309	1.83 (0.82)	0,12
Functioning	0.81 (0.52)	1.42 (0.79)	1.61 (0.71)	62.905	1.47 (0.77)	0,06
Risk	0.07 (0.24)	0.21 (0.38)	0.29 (0.45)	15.385	0.23 (0.40)	1,03
Non-risk items	0.86 (0.59)	1.64 (0.77)	1.81 (0.72)	86.677	1.69 (0.76)	0,15
All items	0.72 (0.50)	1.39 (0.67)	1.58 (0.64)	93.515	0.94 (0.42)	0,65
BDI-II	7.07(7.16)	20.95 (11.56)	25.11(11.84)	138.49	21.99 (11.76)	0,50
BAI	6.90(7.4)	17.16 (12.66)	17.11(11.30)	56.08	17.16 (12.32)	0,55

*All F-values are significant at the $p<0.001$ level.

Table 9. Sensitivity to change.

		First		Last		
	N	Mean	S.D.	Mean	S.D.	Cohens d
Problems	243	1.84	0.81	1.50	0.80	0.42
Function	243	1.96	1.00	1.48	0.96	0.48
Well-being	243	1.45	0.79	1.23	0.75	0.28
Risk	244	0.23	0.38	0.17	0.35	0.16
Non–Risk items	244	1.69	0.76	1.38	0.75	0.41
All items	210	1.41	0.67	1.16	0.66	0.37